



MDD diagnosis based on EEG feature fusion and improved feature selection



Wan Chen, Yanping Cai ^{*}, Aihua Li, Yanzhao Su, Ke Jiang

Rocket Force University of Engineering, Xi'an 710025, China

ARTICLE INFO

Keywords:

EEG
Major depression disorder
Feature selection
Functional connectivity

ABSTRACT

Conventional scale-based major depression disorder (MDD) diagnosis methods are subjective, so it is significant to propose an objective and accurate MDD diagnosis method to assist physicians in diagnosing MDD. This paper proposes an MDD diagnostic method based on electroencephalogram (EEG) feature fusion and improved feature selection. First, seven functional connectivity matrices are extracted and reassembled into a vector to obtain the fusion functional connectivity feature. Then, a feature selection method based on principal component analysis, K-means, and mutual information (PKM) is constructed to optimize the high-dimensional EEG features. Finally, seven classifiers are used for MDD diagnosis. The results show that the proposed method performs better than the existing methods in MDD diagnosis with accuracy, sensitivity, and specificity of 88.73%, 90.67%, and 86%, respectively. Phase lag index (PLI) and phase-locked value (PLV) features, alpha and delta bands contribute significantly to MDD diagnosis. Functional connectivity in the right hemisphere of the brain, particularly in the right temporal and central prefrontal regions with other brain regions, may be beneficial for MDD diagnosis. High-precision MDD diagnosis can be achieved using EEG from only four channel pairs. In summary, this study provides an objective and accurate method for MDD diagnosis.

1. Introduction

Major depression disorder (MDD) is a common mental illness [1]. If the patient is not treated in time, it may cause the patient's condition to worsen and even threaten the patient's life [2]. According to the World Health Organization, more than 300 million people in the world are suffering from depression [3]. Most of the traditional diagnostic methods of depression are judged by experts based on the scale and clinical experience [4]. However, as the results of the scale depending on the cooperation of patients, the judgment results of experts are affected by the level of personal professional knowledge [5], resulting in the high subjectivity and low accuracy of traditional depression diagnosis methods. Therefore, it is of great significance to propose an objective and accurate method for depression diagnosis.

As an objective physiological signal, electroencephalogram (EEG) has been widely used in MDD diagnosis due to its advantages of high time resolution, low cost, easy operation, and non-trauma [6–8]. Feature extraction is critical to MDD diagnosis. Mohammed et al. [9] extracted a variety of linear and nonlinear features from EEG and implemented MDD diagnosis using support vector machine (SVM) and K-nearest neighbor (KNN). Hamid et al. [10] extracted features such as mean value, variance, kurtosis, and Hjorth from EEG and fused facial features

to improve the accuracy of MDD diagnosis. Khadidos et al. [11] extracted power spectral density (PSD) from four sub-bands of EEG. However, the above methods treat the electrodes as isolated points, making extracted EEG features sensitive to signal amplitude [12]. Due to the nonlinearity and complexity of EEG signals, the EEG between different subjects will be significantly different [13]. Therefore, the accuracy and robustness of the above methods in cross-subject MDD diagnosis are not high.

Functional connectivity is a promising EEG feature widely used in MDD diagnosis [14–16]. Yang et al. [17] calculated the functional connectivity features using the Pearson correlation coefficient (PCC), optimized the feature set using the wrapped feature selection method, and finally adopted SVM to realize MDD diagnosis. Li et al. [18] constructed the functional connectivity matrix using coherence (Coh) and found that the Coh of MDD was significantly higher than that of the healthy control (HC). Shim et al. [19] extracted PSD, phase-locked value (PLV), and brain network parameters based on PLV from EEG. The results show that the PLV-based method has the highest accuracy. Besides, various connectivity methods have also been proposed in existing studies [20]. PCC [21] and Coh [22] can measure the correlation between signals from the perspective of the time domain and frequency domain, but these two methods can only detect the linear correlation

* Corresponding author.

E-mail address: [caiyaping502@163.com](mailto:caiyanping502@163.com) (Y. Cai).

between signals. Mutual information (MI) [23] measures the correlation between signals from the information theory perspective and can detect the linear and nonlinear relationships of signals. The PLV [24] and the phase lag index (PLI) [25] measure the correlation between signals from the perspective of phase synchronization. Existing studies have shown that PLV is more sensitive to volume conduction, while PLI is not [26]. There are also connectivity methods that are insensitive to volume conduction and noise, such as imaginary part coherence (ICoh) [27] and weighted phase lag index (WPLI) [28]. Different connectivity methods can measure functional connectivity of MDD from different perspectives. Therefore, the fusion of different features may help to improve the accuracy and robustness of MDD diagnosis. However, most existing studies use single functional connectivity features to diagnose MDD. Meanwhile, few studies have compared the performance of different functional connectivity features in MDD diagnosis.

With the progress of sensor technology, the number of EEG acquisition equipment channels increases, leading to a high EEG feature dimension [29]. In addition, feature fusion will also increase the feature dimension to a certain extent [30]. On the one hand, high-dimensional features lead to high algorithm complexity, and irrelevant and redundant features may reduce the accuracy of the model [31]. On the other hand, multi-channel EEG acquisition equipment is expensive and inconvenient to use, which is not conducive to daily promotion and use [32]. Therefore, feature selection is often used after feature extraction to reduce the feature dimension and improve the classification accuracy [33]. Meanwhile, according to the feature selection results, the importance of channel pairs is analyzed to achieve MDD diagnosis with fewer channels [34].

Feature selection methods can be divided into filter feature selection [35–37] and wrap feature selection [38–40]. The filtering feature selection method is relatively simple and widely used in EEG feature selection [41]. Mohan et al. [42] extracted time domain features, frequency domain features, and nonlinear features from EEG and optimized the feature set using maximum correlation minimum redundancy (MRMR). Tuncer et al. [43] proposed an EEG feature selection method that combines ReliefF and neighbor component analysis (RFINCA), and the results show that the proposed method has good feature optimization capability. Peng et al. [44] extracted the PLI from the 128-channel EEG and optimized the feature set using the feature selection based on the rank correlation coefficient. The distribution of essential channel pairs is given, but the diagnosis of MDD based on fewer channels is not discussed. Wu et al. [45] extracted the power spectrum and relative power from the 29-channel EEG and adopted the Fscore for feature selection. The results show that seven channels can achieve the same classification performance as 29. In addition, correlation-based feature selection (CFS) is also commonly used for EEG feature selection [46]. Due to the differences between environments and subjects, the EEG of different subjects can be pretty different, affecting the performance of feature selection methods. However, existing methods barely consider the sample distribution in EEG samples. Moreover, the application of feature fusion and multi-channel EEG leads to a sharp increase in the dimensionality of EEG features. How to efficiently remove irrelevant features and redundant features from high-dimensional EEG features is still an open problem.

In order to improve the diagnostic accuracy of MDD, more attention should be paid to samples with low separability, that is, samples with close distance between MDD and HC, during feature selection. Therefore, the proposed method extracts subsets with low separability before computing the weights of the features. Clustering is an unsupervised machine-learning method [33]. Clustering algorithms can partition the dataset into several disjoint subsets based on the distance between samples, which can be used to extract subsets with low separability. The K-means clustering algorithm is a standard clustering method, which is simple to calculate and has been widely used in many fields [47–49]. Therefore, the K-means algorithm is used in this paper to extract subsets with low separability. In addition, to avoid the influence of uncorrelated

and redundant features on clustering, the proposed method adopts principal component analysis (PCA) [50–52] to extract the principal components of the dataset before clustering. Then, on the obtained subset, the weights of the features are computed. Fscore and ReliefF measure the importance of features according to the distance between features and categories but ignore the redundancy between features [43,45]. Mutual information (MI) and Pearson correlation coefficient (PCC) can simultaneously measure the correlation between features and categories and the correlation between features and can effectively eliminate irrelevant and redundant features [42,46]. Thus, the proposed method uses MI and PCC to measure the weights of features.

Therefore, this paper proposes an MDD diagnosis method based on fused features and improved feature selection. The contribution of this paper is as follows.

- (1) Seven functional connectivity features were extracted, including Coh, ICoh, MI, PCC, PLI, PLV, and WPLI, and the performance of fusion features and single features in MDD diagnosis is analyzed. The results show that the fused features perform better than the single features in MDD.
- (2) A feature selection method based on PCA, K-means, and mutual information (PKM)/correlation coefficient (PKC) is proposed, and ablation experiments and comparison experiments with existing methods are designed. The results show that PKM outperforms PKC and existing feature selection methods.
- (3) The distribution of feature, frequency band, and channel pairs in feature subsets is explored, and the results are discussed in detail. The results show that PLI and PLV features, as well as delta and alpha bands, contribute significantly to the diagnosis of MDD. Functional connectivity of the brain's right hemisphere, especially the right temporal region (T4 and T6) and the center of the forehead (Fz) with other brain regions, facilitates the diagnosis of MDD.
- (4) The performance of fewer channels on the diagnosis of MDD is explored. The results show that the effect of six channels (four channel pairs) for MDD diagnosis is comparable to that of 19 channels. The results can provide a basis for diagnosing MDD based on fewer channels, which is conducive to applying the method in daily life.

2. Method

Fig. 1 shows the methodology used in this paper, including data preprocessing, feature extraction, feature selection and classification, and the result analysis. The proposed method mainly performs artifact removal, data segmentation, and band extraction in preprocessing. For feature extraction, seven functional connectivity features, including Coh, ICoh, MI, PCC, PLI, PLV, and WPLI, are extracted, and fused features are obtained. The dataset is divided into training and test sets in feature selection and classification. First, the feature weights are calculated using PKM. Then, a forward search algorithm is used to find the feature subset. Finally, seven classifiers are used to achieve MDD classification. In the result analysis, the superiority of the proposed method is verified from various aspects, and MDD diagnosis based on fewer channel pairs is achieved.

2.1. Functional connectivity feature extraction

For EEG signals with N channels, an $N \times N$ functional connectivity matrix can be constructed [14]. The functional connectivity matrix is symmetric, and the diagonal elements represent autocorrelations [17]. Thus, the actual eigenvector consists of the upper triangular elements of the matrix. For an EEG signal with N channels, the dimension of the single eigenvector is $N \times (N-1)/2$. This paper uses seven functional connectivity methods to extract EEG features [21–28], including Coh, ICoh, MI, PCC, PLI, PLV, and WPLI. Detailed information is shown in

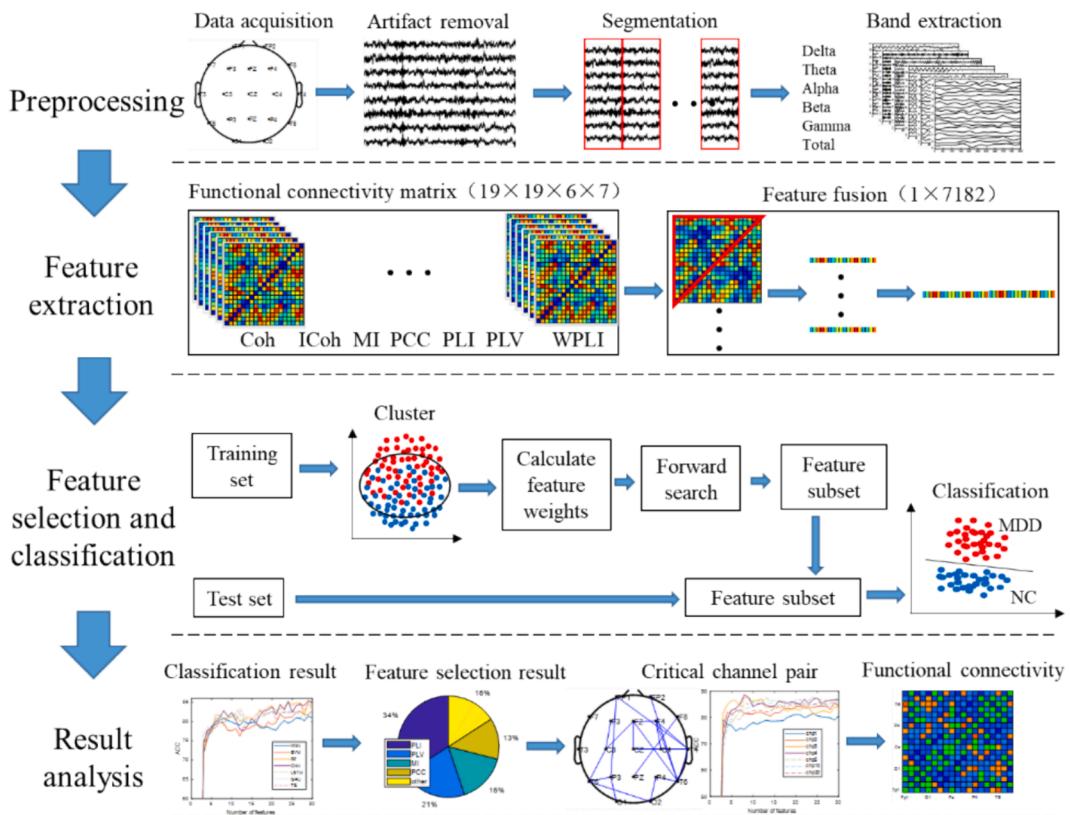


Fig. 1. Flow diagram of the proposed method.

Table 1
Specific calculation methods of feature extraction.

Method	Calculation formula	Remark
Coh	$Coh_{xy}(f) = \frac{ p_{xy}(f) }{\sqrt{p_{xx}(f)p_{yy}(f)}}$	f represents the frequency, $p_{xy}(f)$ represents the mutual PSD of x and y . $p_{xx}(f)$ and $p_{yy}(f)$ represent the PSD of x and y , respectively. $im(\cdot)$ represents the imaginary part, $p(x_i, y_j)$ represents the joint probability distribution function of x_i and y_j . $p(x_i)$ and $p(y_j)$ represent the edge probability distribution of x_i and y_j , respectively. $\varphi_{xy}(t)$ represents the phase difference between x and y at time t , $sign(\cdot)$ represents the sign function, S_{xy} represents the cross-power spectrum of x and y , $ \cdot $ represents the absolute value
ICoh	$ICoh_{xy}(f) = \frac{im(p_{xy}(f))}{\sqrt{p_{xx}(f)p_{yy}(f)}}$	
MI	$MI_{xy} = \sum_{x_i \in x, y_j \in y} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)}$	
PCC	$PCC_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$	
PLI	$PLI_{xy} = \frac{1}{T} \sum_{t=1}^T sign(\varphi_{xy}(t))$	
PLV	$PLV_{xy} = \left \frac{1}{T} \sum_{t=1}^T e^{j\varphi_{xy}(t)} \right $	
WPLI	$WPLI_{xy} = \frac{\left \frac{1}{N} \sum_{n=1}^N im(S_{xy}^n) sign(im(S_{xy}^n)) \right }{\frac{1}{N} \sum_{n=1}^N im(S_{xy}^n) }$	

Table 1.

2.2. Feature selection based on PKM and PKC

In this paper, two feature selection methods are proposed. The first method is based on PCA, K-means, and MI (PKM), and the second is based on PCA, K-means, and correlation coefficient (PKC). PKM and PKC can be divided into two parts: First, PCA and K-means are used to perform cluster analysis on the dataset, eliminating the subsets with high separability and keeping the subsets with low separability. Then, MI and PCC are used to measure the feature weights on the sample

subset to extract feature subsets. For dataset $X_{m \times n} = \{x_1, x_2, \dots, x_m\}$, where m denotes the number of samples and n is the feature dimension, the specific implementation procedure for PKM and PKC is as follows.

(1) Extract sample subsets with low separability

Direct use of K-means for clustering analysis on high-dimensional datasets may lead to errors in the clustering results. Therefore, PCA is used to extract the principal components of the dataset before clustering.

Step1: Calculate the covariance matrix of the sample set X .

$$\bar{X} = \frac{1}{m} \sum_{i=1}^m X_{ij} \quad (1)$$

$$Cov = \frac{1}{m} (X - \bar{X})^T (X - \bar{X}) \quad (2)$$

Step2: Orthogonal decomposition of the covariance matrix.

$$Cov = V \cdot \Sigma \cdot V^T \quad (3)$$

where $\Sigma = diag(\lambda_1, \lambda_2, \dots, \lambda_n)$ represents the diagonal matrix composed of n eigenvalues of the covariance matrix, and $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$. $V = [v_1, v_2, \dots, v_n]$ is a matrix of eigenvectors corresponding to eigenvalues λ_i .

Step3: Calculate the cumulative contribution rate S_a .

$$a_i = \lambda_i / \sum_{i=1}^n \lambda_i, i = 1, 2, \dots, n \quad (4)$$

$$S_a = \sum_{i=1}^n a_i \quad (5)$$

The number of principal components depends on the cumulative contribution rate, which usually ranges from 85 %–95 % [50–52]. In this paper, the contribution rate is set to 90 %. The first s eigenvectors $V_s = [v_1, v_2, \dots, v_s]$ are selected from V , according to the cumulative contribution rate.

Step4: Calculate the principal components of X . $Y_{m \times s} = \{y_1, y_2, \dots, y_m\}$ contains the principal component of the sample set, and s represents the principal component dimension of each sample.

$$Y_{m \times s} = X V_s \quad (6)$$

Step5: Randomly select a sample from $Y_{m \times s}$ as the initial cluster center $c_1 = y_i$, and obtain the cluster center set $c = \{c_1\}$.

Step6: Calculate the distance between each sample point y_i and the existing cluster center c .

$$D_i = \sqrt{\sum_{l=1}^s (y_{il} - c_{jl})^2}, i = 1, 2, \dots, m \quad (7)$$

where D_i represents the distance between y_i and the existing cluster center c . k represents the number of cluster centers.

Step7: Take D_i as the probability that y_i is selected, and use roulette to select the next cluster center c_2 .

Step8: Repeat Step 6–7 until k cluster centers are obtained $c = \{c_1, c_2, \dots, c_k\}$.

Step9: Calculate the distance between the sample y_i and the cluster center c_j .

$$D_{ij} = \sqrt{\sum_{l=1}^s (y_{il} - c_{jl})^2} \quad (8)$$

Step10: Group sample y_i according to D_{ij} . Sample y_i is divided into clusters closest to the cluster center. For k cluster centers, sample $Y_{m \times s}$ is divided into k clusters.

$$y_i \in c_r \quad D_{ir} \leq D_{ij}, j = 1, 2, \dots, k \quad (9)$$

Step11: Recalculate the cluster centers of k clusters.

$$c_i = \frac{1}{n_{c_i}} \sum_{y_i \in c_i} y_i \quad (10)$$

where n_{c_i} represents the number of samples contained by the cluster whose center is c_i .

Step12: Repeat Step9–11 until cluster center c does not change. The cluster center $c = \{c_1, c_2, \dots, c_k\}$ and cluster $\{Y^{c_1}, Y^{c_2}, \dots, Y^{c_k}\}$ of $Y_{m \times s}$ are obtained. According to the clustering results $Y_{m \times s}$, k clusters $\{X^{c_1}, X^{c_2}, \dots, X^{c_k}\}$ of $X_{m \times n}$ are obtained.

Step13: Measure the separability of k clusters. In the MDD dataset, cluster X^{c_i} may contain only one class of MDD or HC, or it may contain both MDD and HC. When X^{c_i} contains almost only one class of samples, the samples in X^{c_i} are far away from the other class of samples and have good separability. When the number of MDD and HC samples in X^{c_i} is similar, it indicates that X^{c_i} is a mixed region of MDD and HC, hence having low separability. From this, the separability measure of a cluster can be defined in terms of the number ratio of MDD and HC in each cluster.

$$E_i = \left| \frac{N_{c_i}^{MDD}}{N_{c_i}} - 0.5 \right|, i = 1, 2, \dots, k \quad (11)$$

where N_{c_i} represents the number of samples contained in the cluster centered at c_i . $N_{c_i}^{MDD}$ represents the number of MDD samples in the cluster. The value of E ranges from 0 to 0.5. The separability index $E = 0.5$ when the cluster contains only MDD or HC. The separability index $E = 0$ when the number of MDD and HC contained in the cluster is equal. Thus, the larger the value of E , the better the separability of the clusters.

Step14: Extract the less separable subset from A. K clusters are sorted in descending order according to E , and clusters with larger E are eliminated. The remaining clusters form a subset with lower separability. Since the MDD diagnosis in this paper is a two-class problem, as the cluster center k increases, at least two clusters with higher E will appear: the cluster containing almost only MDD and the cluster con-

taining almost only HC. Also, if too many clusters are removed, this may result in too few samples. Therefore, the two clusters with higher E are eliminated in this paper, and the subset with the lower separability is obtained.

(2) Extract feature subsets based on feature weights

The weights of the features to be selected are mainly measured in two ways: First, the correlation between features and categories is calculated to measure the useful information contained in the features to be selected. The second is to compute the correlation between the feature to be selected and the feature to measure the redundancy of the feature to be selected. The proposed method measures the feature weights on the subset $Z_{m \times n} = \{f_1, f_2, \dots, f_n\}$. Where f_i denotes i -th feature.

Step15: Calculate the correlation of all features to the class

$$fL_i = G(f_i, L) \quad (12)$$

where $G(\cdot)$ represents the correlation measurement function and L represents the label.

Step16: The feature f_{i1} with the largest fL is the first feature and forms the feature subset $B = \{f_{i1}\}$.

Step17: Calculate the correlation between the remaining features and the class.

$$fL_i = G(f_i, L) \quad (13)$$

Step18: Calculate the correlation between the remaining features and feature subset B .

$$ff_i = \frac{1}{l} \sum_{j=1}^l G(f_i, B_j) \quad (14)$$

Step19: Calculate the weights of the remaining features.

$$W_i = G(f_i, L) - \frac{1}{l} \sum_{j=1}^l G(f_i, B_j) \quad (15)$$

Step20: The feature f_{i2} with the largest W is the next feature and constitutes the feature subset $B = \{f_{i1}, f_{i2}\}$.

Step21: Repeat Step17 to Step20 until the stop condition is met.

PKM

- Input: Training set X , label L , Number of cluster centers k , Threshold s
 - Output: Feature subset B
 - **Extract sample subsets with low separability:**
 - Calculate the covariance matrix of X according to equations (1) – (2).
 - Decompose the covariance matrix according to equation (3), and the eigenvector matrix is obtained $V = [v_1, v_2, \dots, v_n]$.
 - Calculate the cumulative contribution rate S_a according to equations (4) – (5) and obtain the projection matrix $V_s = [v_1, v_2, \dots, v_s]$.
 - Calculate the principal components $Y_{m \times s} = \{y_1, y_2, \dots, y_m\}$ of X according to equation (6).
 - Initialize k cluster centers according to Step5–Step8.
 - Calculate the cluster center $c = \{c_1, c_2, \dots, c_k\}$ of $Y_{m \times s}$ and the cluster $\{X^{c_1}, X^{c_2}, \dots, X^{c_k}\}$ of $X_{m \times n}$ according to Step9–12
 - Calculate the separability index E of each cluster according to equation (11)
 - Remove the two clusters with the largest E from $X_{m \times n}$ and the sample subset $Z_{m \times n}$ with low separability is obtained.
 - **Extract feature subsets based on feature weights:**
 - Calculate the correlation between all features and classes according to equation (12);
 - Add the feature f_{i1} with the largest correlation to the feature subset $B(1) = f_{i1}$.
 - $m = 2$
 - while $m < s$
 - Calculate the correlation of the remaining features and classes according to equation (13).
 - Calculate the correlation between the remaining features and feature subset according to equation (14).
 - $W_i = G(f_i, L) - \frac{1}{l} \sum_{j=1}^l G(f_i, B_j)$
 - Add the feature f_{i2} with the largest W to the feature subset $B(m) = f_{i2}$.
 - $m = m + 1$
 - end while
 - Return B
-

This paper uses MI [42] and PCC [46] to measure the correlation between features and categories and the correlation between features. Therefore, the specific calculation method of $G(\cdot)$ is as follows. Meanwhile, the forward search strategy is used to search the optimal subset [43–45]

$$G(x, y) = \begin{cases} \sum_{x_i \in x, y_j \in y} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(x_i)p(y_j)} & PKM \\ \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} & PKC \end{cases} \quad (16)$$

2.3. Classifier

In order to accurately evaluate the performance of feature subsets, this paper uses seven classifiers to classify MDD, including K-nearest neighbor (KNN) [36], support vector machine (SVM) [15], random forest (RF) [50], convolutional neural network (CNN) [11], Long short-term memory (LSTM) [53], Gated Recurrent Unit (GRU) [53] and Transformer-encoder (TE) [54]. SVM based on the Gaussian kernel function performs well in MDD diagnosis [15], so this paper adopts SVM based on the Gaussian kernel function and the realization of SVM and super parameter optimization using libsvm 3.1 toolkit (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/>).

The structure of CNN is shown in Fig. 2.

The structure of TE is shown in Fig. 3.

3. Data acquisition and preprocessing

3.1. Data acquisition

The data set used in this paper was provided by Mumtaz [55]. 34 MDD patients and 30 HCs participated in the experiment. The detailed information of the subjects is shown in Table 2. The MDD patients were from Hospital Universiti Sains Malaysia (HUSM) and met the internationally recognized diagnostic criteria for MDD, the Diagnostic and Statistical Manual (DSM-IV) [55]. Participants in the MDD group who had other psychiatric disorders, smoking, or alcohol abuse that could affect the collection of EEG were excluded, as were participants in the HC group who had any mental or physical disorders. Before the experiment, all subjects were informed of the contents and signed informed consent. The ethics committee of HUSM approved the experiment.

Five minutes of resting EEG were recorded while the eyes were open. The 19-channel EEG cap was used to collect EEG (Fp1, F3, C3, P3, O1, F7, T3, T5, Fz, Fp2, F4, C4, P4, O2, F8, T4, T6, Cz and Pz), and the electrode placement was based on the international standard 10–20 system. The reference electrode is linked ear (LE) with a sampling frequency of 256Hz. Due to partial deficiencies in the dataset downloaded from the website, the data available included 28 HCs (unavailable for subject 14/25) and 30 MDD patients (unavailable for subject 7/8/12/34).

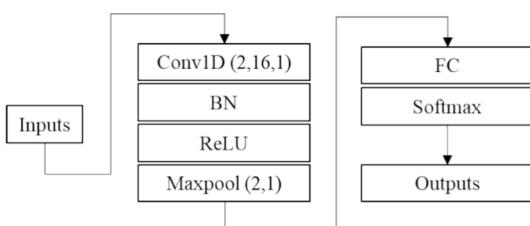


Fig. 2. Structure of CNN.

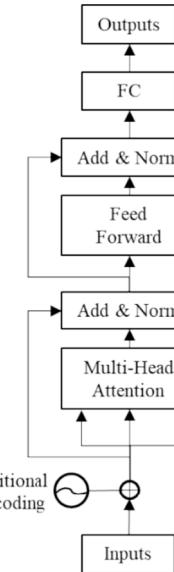


Fig. 3. Structure of Transformer-encoder.

Table 2

The detailed information of the subjects.

Information	MDD	HC
Gender (female/male)	17/17	9/21
Age (years)	40.3 ± 12.9	38.3 ± 15.6
BDI	20.6 ± 8.6	<10

3.2. Data preprocessing

The EEG is easily disturbed by internal and external factors during acquisition, which leads to various noises and artifacts in the EEG. Therefore, before the research, EEG needs to be preprocessed to improve the signal-to-noise ratio of the signal [15]. Data preprocessing was performed on EEGLAB toolkit. The process is as follows:

- (1) Data filtering: 50HZ notch filtering and 0.5–50HZ bandpass filtering are used to remove low-frequency drift, high-frequency noise, and power frequency interference;
- (2) Artifact removal: Wavelet independent component analysis (wICA) [5] is used to remove artifact signals;
- (3) Re-reference: re-reference EEG data to the whole brain average to reduce the distortion of connection mode;
- (4) Data segmentation: To ensure data stability and continuity, 100 s of data after 60 s are extracted for research. The data is split into 10 s to increase the sample size. Finally, each subject can obtain ten pieces of data; the size of each data file is 19*2560 (256*10 s), and the total data size is 58*10 = 580.
- (5) Frequency band extraction: zero-phase digital filter is used to extract six bands, including delta (0.5–4HZ), theta (4–8HZ), alpha (8–14HZ), beta (14–30HZ), gamma (30–50HZ) and one full band (0.5–50HZ).

4. Results

In this section, we validate the effectiveness of the proposed method from multiple aspects and discuss the results in detail. First, the experimental setup is presented. Second, we analyze the performance of PKM and PKC in MDD diagnosis. Third, the specific results of feature selection are analyzed. Fourth, ablation experiments are designed to validate the effectiveness of the improved strategy. Fifth, comparative experiments are designed to verify the superiority of the proposed method. Sixth, we

analyze key channel pairs in MDD diagnosis. Finally, the performance of fewer channel pairs in MDD diagnosis is analyzed. All experiments were performed on MATLAB 2023b and Intel(R) Xeon(R) CPU, 2.3 GHz, 128 GB RAM.

4.1. Experiment details and setup

The number of EEG channels is 19, so the dimension of each functional connectivity matrix is 19×19 . Take the upper triangular elements of the matrix to form the eigenvector, whose dimension is $19 \times (19-1)/2 = 171$. In the preprocessing, six frequency bands are extracted from EEG, so the dimension of a single functional connectivity feature is $171 \times 6 = 1026$. Seven functional connectivity features are extracted and fused, so the dimension of fusion features is $1026 \times 7 = 7182$. The dataset contains 58 subjects, each divided into ten samples. Thus, the total sample size is 580. The dimension of the entire sample matrix is 580×7182 .

To fully utilize the dataset and improve the reliability of the results, 10-fold cross-validation was used in the experiments. The data set was divided into ten pieces, one for the test set and nine for the training data. The training data was divided into the training set and validation set according to 8:2. The Training set and validation set were only used for model training and debugging, and the test set was only used to test model performance. It should be noted that the dataset was divided according to the subjects in the experiment; that is, samples from the same subject can only appear in the training or test set. In addition, the early stop method was used during the training of network models (CNN, LSTM, GRU, and TE). The training stopped when the network model's accuracy on the verification set did not increase after 50 epochs. In order to reduce the feature dimension as much as possible and ensure classification accuracy, this paper only discussed the performance of the first 30 dimensional features in MDD diagnosis.

The evaluation indexes of the model are accuracy, sensitivity, and specificity [28]. The parameter settings of the relevant algorithms in the experiments are shown in [Table 3](#).

$$Acc = \frac{T_p + T_N}{T_p + T_N + F_p + F_N} \quad (17)$$

$$SE = \frac{T_p}{T_p + F_N} \quad (18)$$

$$SP = \frac{T_N}{T_N + F_p} \quad (19)$$

where T_p represents the number of correctly classified MDD. T_N represents the number of correctly classified HC. F_p represents the number of samples that misclassified HC as MDD. F_N represents the number of samples that misclassified MDD as HC.

4.2. Results of MDD diagnosis based on the proposed method

The k value (number of cluster centers) affects the performance of PKM and PKC. Therefore, MDD diagnosis results based on PKM and PKC with different values of k are first analyzed. The results of the seven classifiers are averaged, and the results are shown in [Fig. 4](#). It can be

seen from [Fig. 4](#) (a) that when $k = 8$, the ACC curve is significantly higher than the other curves in the feature dimension of 25–30, which indicates that PKM performs better when $k = 8$. As shown in [Fig. 4](#) (b), when $k = 5$, the ACC curve is significantly higher than other curves when the feature dimension is 20–30, indicating that PKC performs better when $k = 5$. Extract the maximum ACC of PKM and PKC under different k values, and the results are shown in [Fig. 4](#)(c). It can be seen from [Fig. 4](#)(c) that the ACC of PKC reaches the maximum value (82.74 %) when $k = 5$. When $k = 8$, the ACC of PKM reaches the maximum value (83.11 %). When $k = 6, 7, 8, 9, 10$, the ACC of PKM is larger than that of PKC, indicating that PKM is better than PKC for MDD diagnosis. In summary, PKM has better feature selection capability than PKC.

The results show that PKM performs better than PKC, and PKM performs best when $k = 8$. Therefore, the subsequent analysis is performed at the PKM with $k = 8$. [Fig. 5](#) shows the MDD diagnosis results based on fusion features and PKM. It can be seen from [Fig. 4](#) (a) that the MDD classification accuracy of the all classifiers increases with the increase of feature dimension. When the feature dimension increases from 25 to 30, the accuracy does not change significantly, indicating that the MDD diagnostic accuracy does not increase significantly if the feature dimension continues to increase. At the same time, increasing the feature dimension leads to high algorithmic complexity. Therefore, only the results for the first 30 dimensional features are analyzed in this paper. [Fig. 5](#) shows that the accuracy, sensitivity, and specificity curves of KNN are significantly lower than those of the other methods, indicating that KNN performs poorly in MDD diagnosis. The accuracy, sensitivity, and specificity curves of CNN are significantly higher than those of most other classifiers, indicating that CNN has better performance in MDD diagnosis.

[Table 4](#) shows the MDD diagnosis results. As can be seen from Table, the dimension of the feature subset (Sub-dim) is significantly reduced compared to the original feature dimension (Ori-dim). The classification accuracy of RF and CNN is significantly higher than that of other classifiers. However, the difference in sensitivity and specificity of RF is significant, which may lead to unbalanced classification results. Therefore, the MDD diagnosis method based on fusion features, PKM, and CNN performed best. The accuracy, sensitivity, and specificity are 85.5 %, 88.67 %, and 81.33 %, respectively.

4.3. Results of feature selection

The training set is different for each cross-validation, resulting in different results for each feature selection. The first five-dimensional features (a total of 50) in 10-fold cross-validation are extracted, and repeated features are selected for analysis, as shown in [Table 5](#). As shown in [Table 5](#), four features are coming from PLI, and the number of PLI is significantly higher than the other features, indicating that PLI contributes more to the diagnosis of MDD among the seven features. Five of the features are from the alpha band, suggesting that the alpha band contributes more to the diagnosis of MDD among the six bands. The mean and variance of the MDD and HC features are calculated, and an independent sample T-test is performed. The P-values of the features are all smaller than 0.05, indicating that the MDD and HC features are significantly different. The smaller PLI and PLV for MDD and the larger PCC, Coh, and MI for MDD than HC suggest that a single connectivity approach may be flawed when analyzing functional connectivity changes in MDD.

[Table 4](#) shows that when the feature dimension is 29, the MDD diagnosis method based on fusion features, PKM, and CNN achieves the highest MDD classification accuracy. The first 29-dimensional features (290 in total) in 10-fold cross-validation are extracted to analyze further the contribution of different features and frequency bands to the diagnosis of MDD. [Fig. 6](#) shows the number and proportion of different features in the feature subset. As can be seen from [Fig. 6](#), the number of PLI is the largest, accounting for 34 %. The number of PLV ranked second, accounting for 21 %. The number of MI and PCC is almost equal,

Table 3

Algorithm parameter Settings in the experiment.

Algorithms	Parameters
SVM	Hyper-parameters: grid method
KNN	Number of neighbors: 5
RF	Number of trees: 50; Depth of tree: 4
CNN	Epochs: 100; Optimizer: Adam; Initial learning rate: 0.001
LSTM	Epochs: 300; Optimizer: Adam; Initial learning rate: 0.01
GRU	Epochs: 300; Optimizer: Adam; Initial learning rate: 0.01
TE	Epochs: 100; Optimizer: Adam; Initial learning rate: 0.001

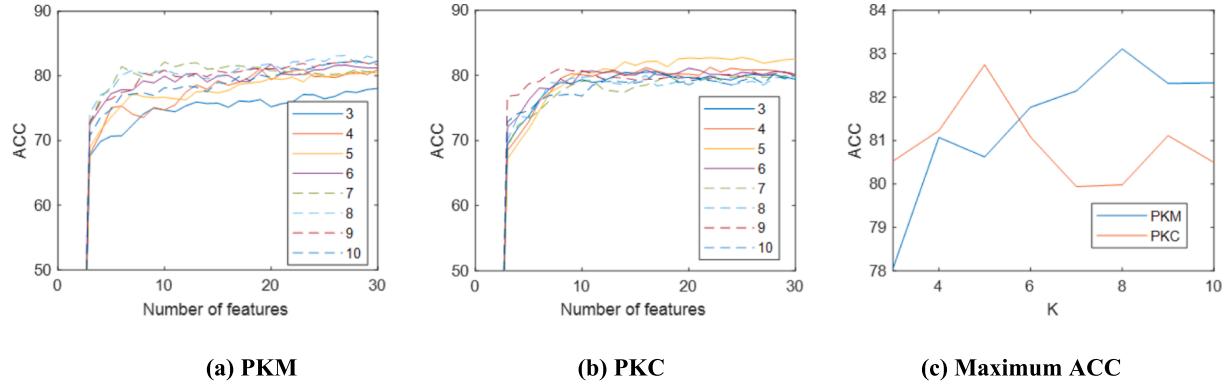
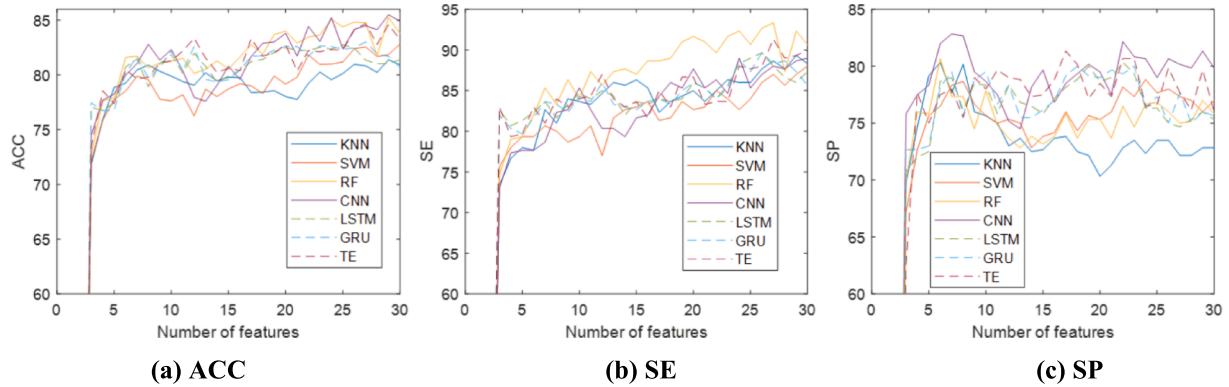
Fig. 4. MDD diagnosis results based on PKM and PKC with different values of k .

Fig. 5. MDD diagnosis results based on PKM.

Table 4
Results of MDD diagnosis based on fusion features and PKM.

Classifier	Ori-dim	Sub-dim	ACC	SE	SP
KNN	7182	29	81.40	89.33	72.83
SVM	7182	30	82.83	88.00	77.00
RF	7182	29	85.27	92.33	77.00
CNN	7182	29	85.50	88.67	81.33
LSTM	7182	25	82.77	89.00	76.33
GRU	7182	27	83.03	88.33	77.67
TE	7182	27	84.70	91.33	77.67

Table 5
Results of feature selection.

ID	Feature	Band	Chp	MDD	HC	P-value
1	PLI	alpha	T5-	0.1618 ±	0.422 ±	<0.05
			C3	0.1016	0.2163	
2	PLI	alpha	O2-	0.6505 ±	0.7892 ±	<0.05
			P4	0.1267	0.0952	
3	PLI	alpha	T6-	0.089 ±	0.1716 ±	<0.05
			P4	0.0632	0.1162	
4	PLI	alpha	T6-	0.456 ±	0.6179 ±	<0.05
			T4	0.1074	0.1043	
5	PCC	delta	T6-Fz	0.643 ±	0.3787 ±	<0.05
				0.1728	0.2087	
6	PCC	gamma	Pz-Fz	0.6478 ±	0.4824 ±	<0.05
				0.1233	0.1705	
7	PLV	beta	Pz-F8	0.1744 ±	0.3119 ±	<0.05
				0.0785	0.1514	
8	Coh	delta	T6-	0.9326 ±	0.8219 ±	<0.05
			P4	0.0484	0.1494	
9	MI	alpha	T6-	0.0534 ±	0.0316 ±	<0.05
			F4	0.0289	0.0143	

accounting for 16 %. The number of Coh, ICoh, and WPLI is small. The results indicate that PLI, PLV, MI, and PCC may be beneficial to the diagnosis of MDD.

Fig. 7 shows the number and proportion of different frequency bands in the feature subset. It can be seen from Fig. 7 that the alpha band has the most significant number, accounting for 46 %. The delta band is second in number, accounting for 32 %. The remaining four frequency bands accounted for a total of 21 %, and their contribution to the diagnosis of MDD was low. The results suggest that alpha and delta bands may benefit MDD diagnosis.

4.4. Ablation experiment

An ablation experiment is performed to verify the effectiveness of the modified strategy. Table 6 details the comparison methods used in the experiments. First, MDD diagnosis is implemented based on the single feature, PKM and CNN. The results are shown in Fig. 8. Fig. 8 shows that PLV have significantly higher accuracy curves than the others.

Method1, Method2, Method3-PLV, Method4 and the proposed method are used for MDD diagnosis, and the results are shown in Table 7. The table shows that the proposed method's accuracy is higher than that of Method1, Method2, and M3-PLV, which indicates that the proposed strategy can improve the method's accuracy. Method4's accuracy is slightly higher than that of the proposed method, but the high-dimensional features are not friendly to practical application. The proposed method uses a small number of EEG features and can achieve classification accuracy comparable to Method4, which shows that the proposed method is effective.

4.5. Comparison between PKM and existing feature selection methods

To validate the superiority of PKM, PKM is compared with existing

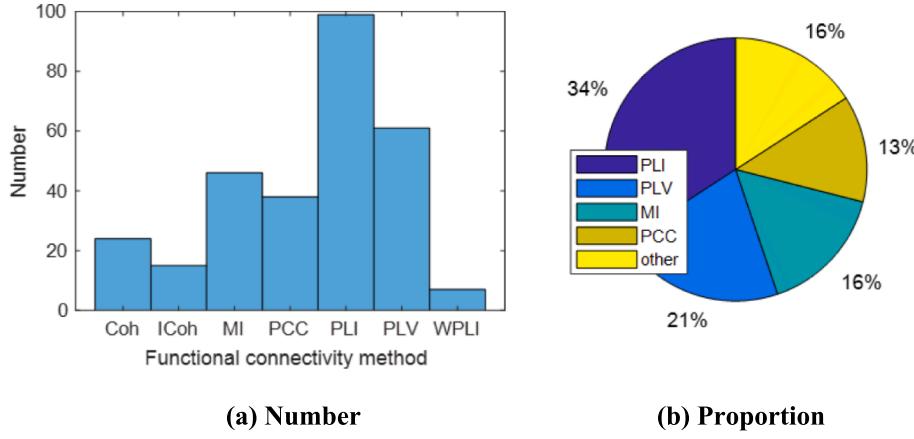


Fig. 6. Distribution of features in feature subset.

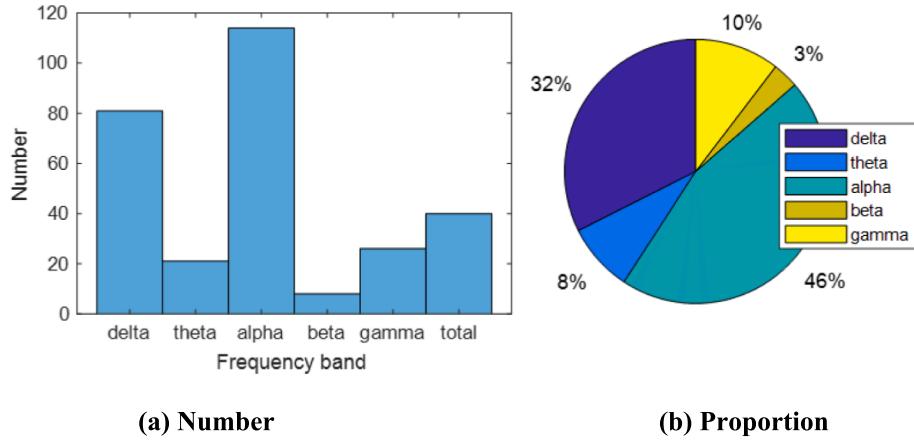


Fig. 7. Distribution of frequency bands in feature subset.

Table 6
Details of the ablation experiment.

Method	Single feature	Fusion feature	PCA	K-means	MI
Method1		✓			✓
Method2		✓			✓
Method3	✓		✓	✓	✓
Method4		✓			
Proposed method	✓	✓	✓	✓	✓

filtering feature selection methods. The comparison algorithms include Fscore [45], MRMR [42], NCA [51], ReliefF [43], RFINCA [43], and CFS [46]. Neighbors of ReliefF and RFINCA set to 6 [43]. Fusion features and CNN are used for MDD diagnosis, and the calculation results of different feature selection algorithms are shown in Fig. 9. As seen in Fig. 9, the accuracy, sensitivity, and specificity curves of the proposed method are higher than those of most other feature selection methods, which indicates that the proposed method performs well in MDD diagnosis.

Table 8 shows the maximum classification accuracy. The classification accuracy and specificity of PKM are 85.50 % and 81.33 %, respectively, which are higher than those of other feature selection algorithms. The sensitivity of PKM is 88.67 %, indicating that it can be used to identify MDD accurately. The results show that PKM can extract effective features from high-dimensional EEG features to achieve high-precision MDD diagnosis and is superior to the existing feature selection algorithms.

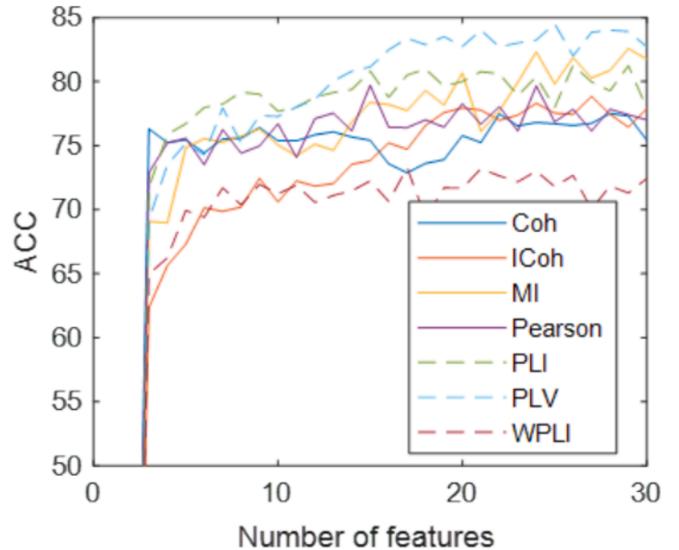


Fig. 8. Results of MDD diagnosis based on single feature and PKM.

4.6. Channel pair distribution in feature subset

To explore the critical channel pairs (Chp) in MDD diagnosis, the frequency of occurrence of 171 channel pairs in the feature subset is analyzed, and the results are shown in Fig. 10. In Fig. 10, the horizontal

Table 7

Comparison of MDD classification methods in ablation experiments.

Method	Ori-dim	Sub-dim	ACC	SE	SP
Method1	7182	27	84.80	88.33	80.33
Method2	7182	28	85.00	89.00	80.00
M3-PLV	1026	25	84.50	85.33	83.33
Method4	7182	\	85.73	88.67	81.83
Proposed method	7182	29	85.50	88.67	81.33

coordinate represents the sequence number of 171 channel pairs, and the vertical coordinate represents the number of channel pairs appearing in the feature subset. The more times a channel pair appears in the feature subset, the more significant the channel pair contribution to the MDD diagnosis and the more helpful information the channel pair contains. Channel pairs are sorted in descending order of their frequency. [Table 9](#) shows the top nine channel pairs.

[Fig. 11](#) shows the brain topology for the first 10, 20, and 30 channel pairs. [Fig. 11](#) shows that for Chp = 10, 2 channel pairs are in the left hemisphere, and seven are in the right hemisphere. T6 has the most significant connections to other electrodes (Number: 4). For Chp = 20, 6 channel pairs are in the left hemisphere, 11 are in the right hemisphere, and 2 channel pairs connect the left and right hemispheres. T6 has the most significant connections to other electrodes (Number: 5). When Chp = 30, 9 channel pairs are in the left hemisphere, 16 channel pairs are in the right hemisphere, and 3 channel pairs connect the left and right hemispheres. Fz has the highest connections to other electrodes (Number 7). Critical channel pairs are more distributed in the right hemisphere, suggesting that brain activity in the right hemisphere may be more favorable for MDD diagnosis. The right temporal region (T4 and T6) and Fz had more connections to other electrodes, suggesting that functional connections between the right temporal/frontal center and other brain regions may be beneficial for MDD diagnosis.

4.7. MDD diagnosis with different number of channel pairs

The more channels an EEG acquisition device has, the higher the cost of the device, which is not conducive to daily use. Therefore, this paper further explores the performance of fewer channels in MDD diagnosis. Based on the sequence of channel pairs, EEG with different numbers of channels (N-Ch) is used for MDD diagnosis. The results are shown in [Fig. 12](#). The figure shows that when chp = 1, the accuracy, sensitivity, and specificity curves are lowest, which indicates that the number of channels is too small and that EEG signals cannot extract enough useful information for MDD classification. When chp = 3/4/5, the accuracy, sensitivity, and specificity curves are higher than those of other methods, indicating that high-precision MDD classification can be achieved with only a small number of channels.

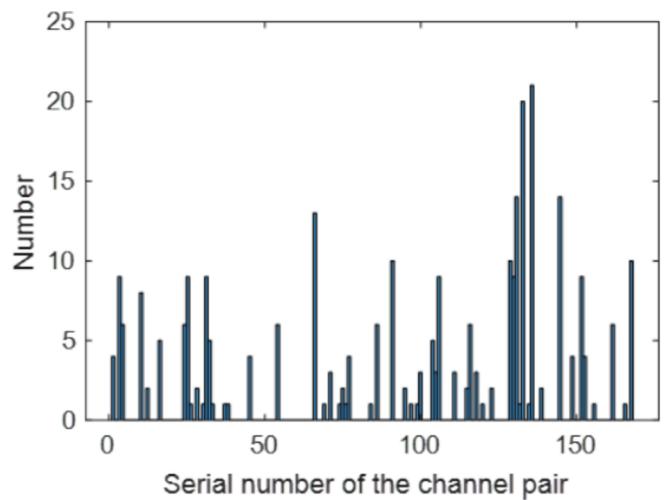
[Table 10](#) shows the maximum MDD classification results. It can be

seen from the table that when 'chp' increases from 1 to 4, the MDD classification accuracy gradually increases, which indicates that EEG can obtain more information conducive to MDD classification with the increase in the number of channels. When 'chp' continued to increase from 4, the MDD classification accuracy gradually decreased, indicating that the redundant information obtained gradually increased with the increase of the number of channels. When chp = 4 (T6-T4, T6-P4, T6-F4, Cz-Fz, channel number: 6), the method performs best, and the accuracy,

Table 8

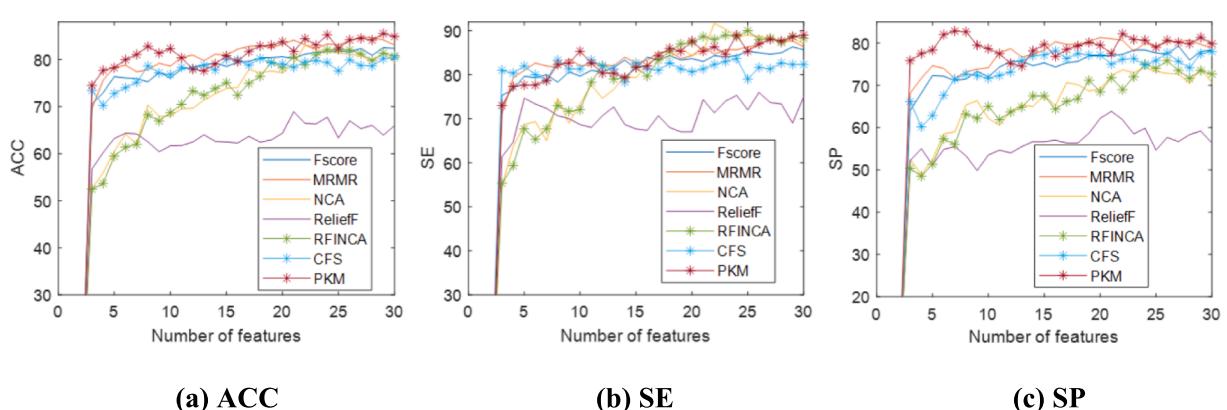
Comparison between PKM and existing feature selection methods.

Methods	Ori-dim	Sub-dim	ACC	SE	SP
Fscore	7182	29	82.53	86.33	78.00
MRMR	7182	27	84.80	88.33	80.33
NCA	7182	22	83.07	91.67	73.67
ReliefF	7182	21	68.90	74.33	63.83
RFINCA	7182	25	82.13	90.00	74.00
CFS	7182	30	80.47	82.33	78.00
PKM	7182	29	85.50	88.67	81.33

[Fig. 10](#). Distribution of channel pairs in feature subsets.**Table 9**

First nine channel pairs.

ID	Chp	ID	Chp	ID	Chp
1	T6-T4	4	Cz-Fz	7	T6-Fz
2	T6-P4	5	C4-F4	8	Pz-F8
3	T6-F4	6	O2-P4	9	C3-F3

[Fig. 9](#). Classification results of different feature selection methods.

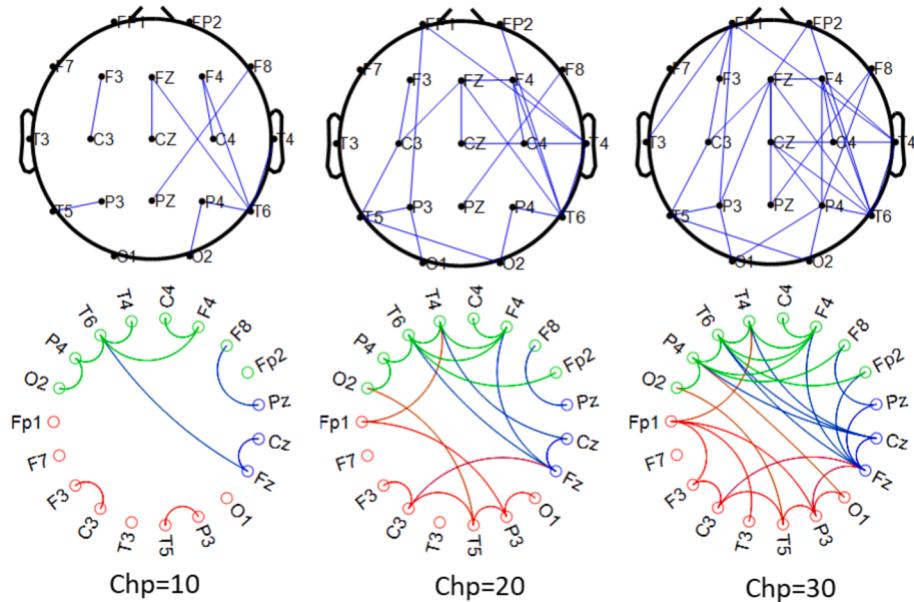


Fig. 11. Brain topology of the first 10/20/30 channel pairs.

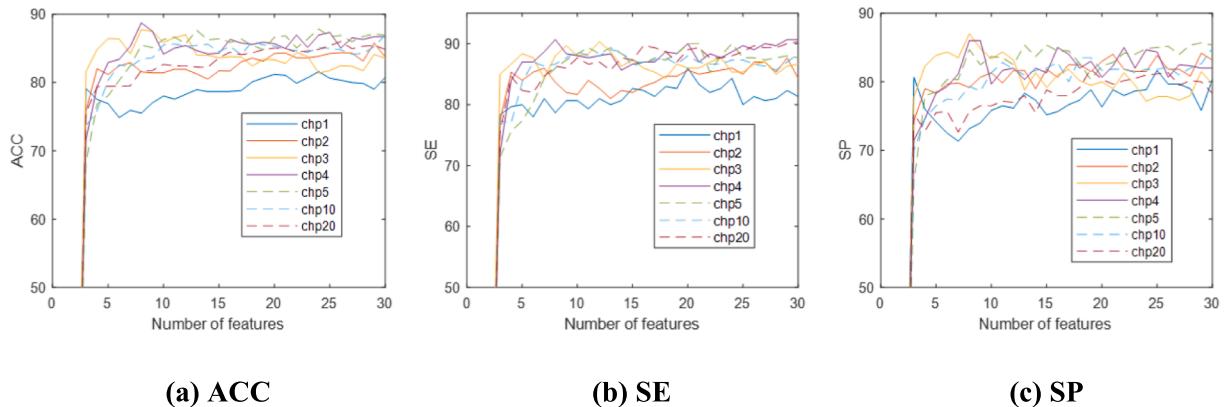


Fig. 12. Classification results based on different number of channel pairs.

Table 10
Results of different number of channel pairs in MDD diagnosis.

N-Chp	N-Ch	Ori-dim	Sub-dim	ACC	SE	SP
1	2	42	24	81.60	84.33	78.83
2	3	84	29	85.77	87.67	84.17
3	4	126	8	87.73	88.00	87.00
4	6	168	8	88.73	90.67	86.00
5	7	210	24	87.83	90.00	84.83
10	14	420	30	86.93	88.33	84.83
20	17	840	26	86.00	89.67	81.33
171	19	7182	29	85.50	88.67	81.33

sensitivity, and specificity are 88.73 %, 90.67 % and 86 %, respectively. Meanwhile, compared with 19 channels, the accuracy of MDD classification using six channels is increased by 3.23 %, the sensitivity is increased by 2 %, and the specificity is increased by 4.67 %, which indicates that using a small number of channels is conducive to MDD classification and can reduce the complexity of signal acquisition.

4.8. Brain functional connectivity changes in MDD

The functional connectivity matrix of MDD and HC is averaged. The mean matrix of HC is then subtracted from the mean matrix of MDD to

obtain the difference matrix, which can be used to analyze changes in the functional connectivity of MDD. If the elements of the difference matrix are larger than zero, the connectivity of MDD is larger than that of HC. Independent sample *t*-test ($P < 0.05$) was used to analyze the significance of MDD and HC, and the channel pairs with significant differences were given. The alpha band is favorable for MDD diagnosis. Therefore, in this paper, we analyze the difference matrix in the alpha band, as shown in Fig. 13. In Fig. 13, red means that the element is greater than zero, blue means that the element is less than zero, and green means that MDD and HC do not differ significantly. In the difference matrix corresponding to Coh, MI, and PCC, the number of red squares is significantly larger than that of blue squares, indicating that the connectivity of the MDD brain regions may have an upward trend. In the difference matrix corresponding to ICoh, PLI, PLV, and WPLI, the number of red squares is significantly lower than the number of blue squares, suggesting that there may be a downward trend in the connectivity of the MDD brain regions. The results suggest that different functional connectivity measures may lead to different results. Therefore, when analyzing changes in brain functional connectivity, using multiple functional connectivity matrices may improve the reliability of the results.

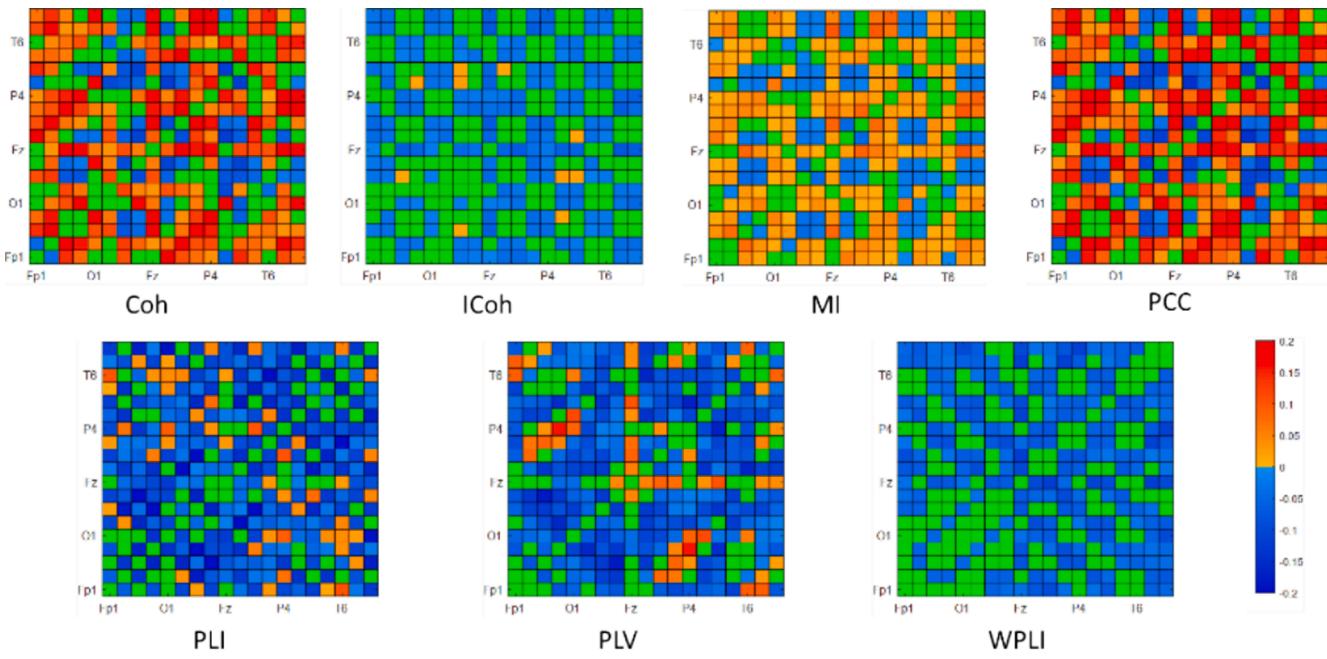


Fig. 13. Functional connectivity changes of MDD in alpha band.

5. Discussion

This paper proposes a method for MDD diagnosis under high-dimensional EEG features based on fusion features, PKM and CNN. The results show that PKM performs better than existing feature selection methods, and fusion features perform better than single-functional connectivity features. The accuracy, sensitivity, and specificity of the proposed method are 85.5 %, 88.67 %, and 81.33 %, respectively. Feature selection results show that PLI and PLV features, as well as alpha and delta frequency bands, significantly contribute to MDD diagnosis. Functional connectivity of the brain's right hemisphere, especially the right temporal (T4 and T6) and the center of the forehead (Fz) with other brain regions, may be beneficial for MDD diagnosis. The results of different numbers of channel pairs in MDD diagnosis showed that using four-channel pairs (6 channels) can achieve the best MDD diagnosis; the accuracy, sensitivity, and specificity are 88.73 %, 90.67 %, and 86 %, respectively. Finally, we analyze the functional connectivity changes in the alpha band. The results show that measures of functional connectivity affect the outcome of brain functional connectivity. Results from ICoh, PLI, PLV, and WPLI showed that the functional connectivity in MDD was lower than that in HC, while results from Coh, MI, and PCC showed that the functional connectivity in MDD was higher than that in HC. In addition, KNN, SVM, RF, CNN, LSTM, GRU and TE are used for MDD classification, and the results show that the accuracy and robustness of CNN in MDD diagnosis are higher than those of the other classifiers.

To verify the superiority of the proposed method, we analyze the performance of the proposed method from several aspects. Firstly, the influence of the k value (number of cluster centers) on PKM and PKC is explored. The results show that PKM performs better than PKC in MDD diagnosis. For $k = 8$, the MDD diagnostic methods based on fused features, PKM, and CNN performed best, achieving accuracy, sensitivity, and specificity of 85.5 %, 88.67 %, and 81.33 %, respectively. Then, the effect of the improved strategy on the proposed method is explored in ablation experiments. Comparing the performance of single features and fusion features in MDD diagnosis, it is found that fusion features perform better than single features. PLV performs better than other features in MDD diagnosis. Comparing the performance of PKM, K-means + MI, and MI in MDD diagnosis, it is found that PKM performs better than the other

two feature selection methods. Finally, PKM is compared with Fscore, MRMR, NCA, Relieff, RFINCA, and CFS. The results show that PKM achieves higher accuracy, sensitivity, and specificity than existing feature selection methods. The effectiveness and superiority of the proposed method are fully validated using ablation experiments and comparison experiments.

The feature selection results are analyzed to explore the impact of different features, frequency bands, and channels on MDD diagnosis. The occurrence times of different features in the feature subset are counted. The results show that PLI and PLV appear more frequently than other features, accounting for 34 % and 21 %, respectively. MI and PCC frequencies are in the middle, at 16 %. Combined with the classification results of individual features, PLV perform well for MDD diagnosis, while Coh, ICoh, and WPLI perform poorly. Existing studies rarely compare the performance of different functional connectivity methods in MDD diagnosis [25], and the results of this study provide a reference for exploring MDD diagnosis methods based on a single feature. The occurrence times of different bands in the feature subsets are counted. The results show that the alpha band appears most frequently, accounting for 46 %; The delta band is the second most frequent, accounting for 32 %, and the other bands are less frequent, indicating that the alpha and delta bands contribute more to the diagnosis of MDD. The existing research results also show that the EEG features of alpha and delta bands perform well in diagnosing MDD [15,16,45,53]. The frequencies of different channel pairs in the feature subsets are counted, the channel pairs are sorted in descending order based on the statistical results, and the importance of the channel pairs is ranked. We explore the brain topology for the first 10, 20, and 30 channel pairs. The results show that most of the channel pairs were distributed in the brain's right hemisphere, and the right temporal region (T4 and T6) and the center of the forehead (Fz) had more connections with other brain regions. Sun et al. [15] used ICoh and graph theory to explore the brain functional connectivity network of MDD, and the results showed that the right hemisphere of MDD was dysfunctional. Li et al. [22] explored the brain network of MDD with Coh, and the results showed abnormal network topologies in the frontal and occipital lobes of MDD. Zhang et al. [25] explored the brain network of MDD using PLI, and the results showed that significant brain synchronous changes occurred in the temporal lobe of MDD. Based on existing studies and the results of this study, the

functional connectivity of the right hemisphere of the brain, especially the functional connectivity of the right temporal region and the center of the forehead with other brain areas, may be conducive to the diagnosis of MDD [56]. Depending on the sequencing results of the channel pairs, we use different numbers of channel pairs for MDD diagnosis. The results showed that when four-channel pairs (T6-T4, T6-P4, T6-F4, Cz-Fz, channel number: 6) were used to diagnose MDD, the classification accuracy was the highest, and the accuracy, sensitivity, and specificity were 88.73 %, 90.67 %, and 86 %, respectively. Compared with the classification method based on all channels EEG, the accuracy of the classification method based on four-channel pairs (six channels) is increased by 3.23 %, the sensitivity is increased by 2 %, and the specificity is increased by 4.67 %. At the same time, the required number of channels is small, which can reduce the price and operation time of the acquisition equipment and is conducive to the application of the method in daily life [28].

The functional connectivity matrix of MDD and NC is used to calculate the difference matrix and explore the changes in brain functional connectivity in MDD. The difference matrix of ICoh, PLI, PLV, and WPLI in the alpha band shows that MDD has lower functional connectivity than HC in most channel pairs. The difference matrix of Coh, MI, and PCC in the alpha band shows that the functional connectivity of MDD is higher than that of HC in most of the channel pairs. Different connectivity methods lead to different results on the same dataset, suggesting that functional connectivity measures significantly influence the construction of functional connectivity networks in the MDD brain. Li et al. [18] used Coh to build functional connectivity networks and found that the Coh of MDD was higher than that of HC. Duan et al. [57] analyzed the brain functional connectivity of MDD using PCC, and the results showed that the PCC of MDD in the alpha band was greater than that of HC. Li et al. [58] realized MDD diagnosis using PSD, fuzzy entropy, and PLI, and the results showed that in the alpha band, the PLI of most channel pairs was smaller than HC. Other studies have used graph theory to explore changes in brain functional connectivity [22]. Some research results show that the path length of MDD increases [59], while others show that the path length of MDD decreases [15]. Both the results of this study and the existing studies suggest that functional connectivity in MDD is affected by connectivity measures and that different findings may emerge. Therefore, when exploring the functional connectivity of the MDD brain, statistical analysis using various connectivity methods can improve the reliability of the results to a certain extent.

Table 11 presents the results of some existing studies. Compared with the existing research, the classification accuracy of the proposed method has a certain competitiveness and is superior to most existing methods. The classification accuracy of the studies [8,50] is slightly higher than that of the proposed method. However, the above methods use 64-channel EEG, which is not conducive to the practical application. Studies [7,12] use 3-channel EEG, but the classification accuracy is slightly lower than that of the proposed method. The proposed method uses only 4-channel EEG, and the classification accuracy reaches 88.73 %, which benefits the practical application.

There are limitations to this study as well. The results show that CNN is highly accurate and robust in MDD diagnosis. However, the CNN designed in this paper is relatively simple. Some CNN and graph convolutional neural networks with more complex structures show better performance in MDD diagnosis [12]. In future work, CNN can be improved to further improve the accuracy of the method. In addition, only the difference matrix is used to analyze the functional connectivity of the MDD brain. Existing studies have shown that the MDD brain network analysis method based on graph theory and functional connectivity can more comprehensively describe the topological changes of the brain [20]. Future research should focus on applying graph theory to MDD brain network analysis.

Table 11
Comparison between the proposed method and existing studies.

Study	Subjects	Channel	Feature	Classifier	Accuracy
Shen [7]	170	3	EMD-based features	SVM	83.27 %
Liu [8]	56	64	PSD, LZC, DFA	SVM	89.29 %
Cai [12]	178	3	linear and nonlinear features	KNN	86.98 %
Sun [15]	32	128	Coherence + Graph theory	SVM	87.50 %
Liu [50]	39	64	PLI + Graph theory	SVM	89.70 %
Mumtaz [55]	64	19	Coherence, P300	LR	87.50 %
Proposed method	58	4	Fusion features	CNN	88.73 %

6. Conclusion

In this paper, we propose an approach for MDD diagnosis based on the fusion feature, PKM and CNN to improve the accuracy of MDD diagnosis under high-dimensional EEG features. Ablation and comparison experiments are designed, and the experimental results show that the proposed method outperforms the existing methods. The classification accuracy, sensitivity, and specificity of the proposed method are 85.5 %, 88.67 %, and 81.33 %, respectively, on the 19-channel EEG dataset. Feature selection results show that PLI and PLV features, as well as alpha and delta bands, significantly contribute to MDD diagnosis. Functional connectivity of the brain's right hemisphere, especially the right temporal region (T4 and T6) and the centre of the forehead (Fz) with other brain regions, facilitates the diagnosis of MDD. The results of MDD diagnosis based on fewer channels showed that four-channel pairs (T6-T4, T6-P4, T6-F4, Cz-Fz, channel number: 6) had the best performance in MDD diagnosis, and the accuracy, sensitivity, and specificity were 88.73 %, 90.67 %, and 86 %, respectively. Finally, the functional connectivity of MDD in the alpha band is analyzed. The results showed that the connectivity method significantly influences the results of functional connectivity construction in MDD.

CRediT authorship contribution statement

Wan Chen: Writing – original draft, Visualization, Validation, Methodology, Investigation, Data curation. **Yanping Cai:** Writing – review & editing, Methodology, Conceptualization. **Aihua Li:** Writing – review & editing, Visualization, Validation. **Yanzhao Su:** Writing – review & editing, Validation. **Ke Jiang:** Writing – review & editing, Validation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

- [1] S. Hashempour, et al., Continuous scoring of depression from EEG signals via a hybrid of convolutional neural networks, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 176–183.
- [2] G. Sharma, et al., DepCap: A smart healthcare framework for EEG based depression detection using time-frequency response and deep neural network, *IEEE Access* 11 (2023) 52327–52338.

- [3] G. Sharma, et al., DepHNN: A novel hybrid neural network for electroencephalogram (EEG)-based screening of depression, *Biomed. Signal Process. Control* 66 (2021) 102393.
- [4] D.M. Khan, et al., Development of wavelet coherence EEG as a biomarker for diagnosis of major depressive disorder, *IEEE Sens. J.* 22 (5) (2022) 4315–4325.
- [5] M. Ravan, et al., Discriminating between bipolar and major depressive disorder using a machine learning approach and resting-state EEG data, *Clin. Neurophysiol.* 146 (2023) 30–39.
- [6] J. Li, et al., Effective connectivity based EEG revealing the inhibitory deficits for distracting stimuli in major depression disorders, *IEEE Trans. Affect. Comput.* 14 (1) (2023) 694–705.
- [7] J. Shen, et al., An improved empirical mode decomposition of electroencephalogram signals for depression detection, *IEEE Trans. Affect. Comput.* 13 (1) (2022) 262–271.
- [8] S. Liu, et al., Alterations in patients with first-episode depression in the eyes-open and eyes-closed conditions: a resting-state EEG study, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 1019–1029.
- [9] H. Mohammed, et al., Improving EEG major depression disorder classification using FBSE coupled with domain adaptation method based machine learning algorithms, *Biomed. Signal Process. Control* 85 (2023) 104923.
- [10] D. Hamid, et al., Integration of deep learning for improved diagnosis of depression using EEG and facial features, *Mater. Today Proc.* 80 (2023) 1965–1969.
- [11] A. Khadidos, et al., Machine learning and electroencephalogram signal based diagnosis of depression, *Neurosci. Lett.* 809 (2023) 137313.
- [12] H. Cai, et al., Feature-level fusion approaches based on multimodal EEG data for depression recognition, *Inf. Fusion* 59 (2020) 127–138.
- [13] S. Soni, et al., Electroencephalography signals-based sparse networks integration using a fuzzy ensemble technique for depression detection, *Biomed. Signal Process. Control* 85 (2023) 104873.
- [14] X. Li, et al., A deep learning approach for mild depression recognition based on functional connectivity using electroencephalography, *Front. Neurosci.* 14 (2020) 192.
- [15] S. Sun, et al., Graph theory analysis of functional connectivity in major depression disorder with high-density resting state EEG data, *IEEE Trans. Neural Syst. Rehabil. Eng.* 27 (3) (2019) 429–439.
- [16] M. Xia, et al., An end-to-end deep learning model for EEG-based major depressive disorder classification, *IEEE Access* 11 (2023) 41337–41347.
- [17] L. Yang, et al., Automatic feature learning model combining functional connectivity network and graph regularization for depression detection, *Biomed. Signal Process. Control* 82 (2023) 104520.
- [18] X. Li, et al., A resting-state brain functional network study in MDD based on minimum spanning tree analysis and the hierarchical clustering, *Complexity* 2017 (2017) 9514369.
- [19] M. Shim, et al., Toward practical machine-learning-based diagnosis for drug-naïve women with major depressive disorder using EEG channel reduction approach, *J. Affect. Disord.* 338 (2023) 199–206.
- [20] L. Ismail L, et al., A graph theory-based modeling of functional brain connectivity based on EEG: a systematic review in the context of neuroergonomics, *IEEE Access* 8 (2020) 155103–155135.
- [21] J. Zhu, et al., EEG based depression recognition using improved graph convolutional neural network, *Comput. Biol. Med.* 148 (2022) 105815.
- [22] Y. Li, et al., Abnormal functional connectivity of EEG gamma band in patients with depression during emotional face processing, *Clin. Neurophysiol.* 126 (11) (2015) 2078–2089.
- [23] W. Mumtaz, et al., A machine learning framework involving EEG-based functional connectivity to diagnose major depressive disorder, *Med. Biol. Eng. Compu.* 56 (2) (2018) 233–246.
- [24] Y. Peng, et al., Electroencephalographic network topologies predict antidepressant responses in patients with major depressive disorder, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 2577–2588.
- [25] B. Zhang, et al., Brain functional networks based on resting-state EEG data for major depressive disorder analysis and classification, *IEEE Trans. Neural Syst. Rehabil. Eng.* 29 (2021) 215–229.
- [26] X. Li, et al., A novel index of functional connectivity: phase lag based on Wilcoxon signed rank test, *Cogn. Neurodyn.* 15 (4) (2020) 621–636.
- [27] S. Sun, et al., Abnormal brain topological structure of mild depression during visual search processing based on EEG signals, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 1705–1715.
- [28] Y. Zhang, et al., Minimal EEG channel selection for depression detection with connectivity features during sleep, *Comput. Biol. Med.* 147 (2022) 105690.
- [29] O. Almanza-Conejo, et al., A channel selection method to find the role of the amygdala in emotion recognition avoiding conflict learning in EEG signals, *Eng. Appl. Artif. Intel.* 126 (2023) 106971.
- [30] X. Xu, et al., EEG feature selection via global redundancy minimization for emotion recognition, *IEEE Trans. Affect. Comput.* 14 (1) (2023) 421–435.
- [31] Y. Wang, et al., An adaptive driver fatigue classification framework using EEG and attention-based hybrid neural network with individual feature subsets, *Biomed. Signal Process. Control* 85 (2023) 105045.
- [32] K. Chen, et al., A novel caps-EEGNet combined with channel selection for EEG-based emotion recognition, *Biomed. Signal Process. Control* 86 (2023) 105312.
- [33] S. Sun, et al., Clustering-fusion feature selection method in identifying major depressive disorder based on resting state EEG signals, *IEEE J. Biomed. Health Inform.* 27 (7) (2023) 3152–3156.
- [34] F. Hasanzadeh, et al., Single channel EEG classification: a case study on prediction of major depressive disorder treatment outcome, *IEEE Access* 9 (2021) 3417–3427.
- [35] H. Aldawsari, et al., Optimizing 1D-CNN-based emotion recognition process through channel and feature selection from EEG signals, *Diagnostics* 13 (2023) 2624.
- [36] H. Cai, et al., A pervasive approach to EEG-based depression detection, *Complexity* 2018 (2018) 5238028.
- [37] B. Zhang, et al., Computer-aided recognition based on decision-level multimodal fusion for depression, *IEEE J. Biomed. Health Inform.* 26 (7) (2022) 3466–3477.
- [38] J. Yedukondalu, et al., Cognitive load detection using Ci-SSA for EEG signal decomposition and nature-inspired feature selection, *Turk. J. Electr. Eng. Comput. Sci.* 31 (5) (2023) 771–791.
- [39] M. Ramesh, et al., EEG channel selection in CIT: a binary Harris Hawks optimization using atime-varying transfer function, *Soft. Comput.* 27 (2023) 11013–11026.
- [40] N. Kouka, et al., EEG channel selection-based binary particle swarm optimization with recurrent convolutional autoencoder for emotion recognition, *Biomed. Signal Process. Control* 8 (2023) 104783.
- [41] A. Aderinwale, et al., Two-channel EEG based diagnosis of panic disorder and major depressive disorder using machine learning and non-linear dynamical methods, *Psychiatry Res. Neuroimaging* 332 (2023) 111641.
- [42] R. Mohan, et al., Classification and detection of cognitive disorders like depression and anxiety utilizing deep convolutional neural network (CNN) centered on EEG signal, *Traitement Du Signal* 40 (3) (2023) 971–979.
- [43] T. Tuncer, et al., EEG-based driving fatigue detection using multilevel feature extraction and iterative hybrid feature selection, *Biomed. Signal Process. Control* 68 (2021) 102591.
- [44] H. Peng, et al., Multivariate pattern analysis of EEG-based functional connectivity: a study on the identification of depression, *IEEE Access* 7 (2019) 92630–92641.
- [45] C. Wu, et al., Depression detection using relative EEG power induced by emotionally positive images and a conformal kernel support vector machine, *Appl. Sci.* 8 (8) (2018) 1244.
- [46] Y. Li, et al., A study on feature selection using multi-domain feature extraction for automated k-complex detection, *Front. Neurosci.* 17 (2023) 1224784.
- [47] H. Ding, et al., Resting-state electroencephalogram relevance state recognition of Parkinson's disease based on dynamic weighted symbolic mutual information and k-means clustering, *J. Biomed. Eng.* 40 (1) (2023) 20–26.
- [48] Y. Wang, et al., Synchronous GPS spoofing identification based on K-means clustering, *J. Electron. Inf. Technol.* 45 (11) (2023) 4137–4149.
- [49] X. He, et al., Determination of the optimal number of clusters in K-means algorithm, *J. Univ. Electron. Sci. Technol. China* 51 (06) (2022) 904–912.
- [50] W. Liu, et al., Functional connectivity of major depression disorder using ongoing EEG during music perception, *Clin. Neurophysiol.* 131 (2020) 2413–2422.
- [51] M.M. Moussa, et al., Explainable computer-aided detection of obstructive sleep apnea and depression, *IEEE Access* 10 (2022) 110916–110933.
- [52] T. Monteiro, et al., Using EEG for mental fatigue assessment: a comprehensive look into the current state of the art, *IEEE Trans. Hum.-Mach. Syst.* 49 (6) (2019) 599–610.
- [53] L. Yang, Y. Wang, X. Zhu, et al., A gated temporal-separable attention network for EEG-based depression recognition, *Comput. Biol. Med.* 157 (2023) 106782.
- [54] B. Halder, T. Anjum, M. Bhuiyan, An attention-based multi-resolution deep learning model for automatic A-phase detection of cyclic alternating pattern in sleep using single-channel EEG, *Biomed. Signal Process. Control* 83 (2023) 104730.
- [55] W. Mumtaz, et al., A wavelet-based technique to predict treatment outcome for major depressive disorder, *PLoS One* 12 (2) (2017) 0171409.
- [56] S. Liao, et al., Major depression detection from EEG signals using kernel eigen-filter-bank common spatial patterns, *Sensors* 17 (6) (2017) 1385.
- [57] L. Duan, et al., Machine learning approaches for MDD detection and emotion decoding using EEG signals, *Front. Hum. Neurosci.* 14 (2020) 284.
- [58] G. Li, et al., Machine learning techniques reveal aberrated multidimensional EEG characteristics in patients with depression, *Brain Sci.* 13 (2023) 384.
- [59] M. Shim, et al., Altered cortical functional network in major depressive disorder: a resting-state electroencephalogram study, *NeuroImage: Clinical* 18 (2018) 1000–1007.